

## CHALLENGES AND FUTURE DIRECTIONS FOR DATA MANAGEMENT IN THE GEOSCIENCES

DOUGLAS C. SCHUSTER, MATTHEW S. MAYERNIK, CHUNG-YI HOU, GREG STOSSMEISTER, ROBERT R. DOWNS, DANIE KINKADE, TRAN B. NGUYEN, MOHAN RAMAMURTHY, AND FUQING ZHANG

The open availability and wide accessibility of digital scientific resources, such as articles and datasets, is becoming the norm for twenty-first-century science. Geoscience researchers are now being asked by funding agencies and scientific publishers to archive and cite data to support open access but often struggle to understand, interpret, and fulfill these requirements. To fulfill the promise of new open data initiatives, 1) scientific resources (e.g., data and software) must be collected and documented properly; 2) repository services, including preservation and storage capabilities, must be maintained,

### GEOSCIENCE DIGITAL DATA RESOURCE AND REPOSITORY SERVICE (GEODARRS) WORKSHOP

**WHAT:** More than 60 individuals from multiple stakeholder groups in data management met to discuss challenges and opportunities within the geosciences.

**WHERE:** Boulder, Colorado

**WHEN:** 7–9 August 2018

supported, and improved over time; and 3) governance institutions must be established.

These issues were discussed in the Geoscience Digital Data Resource and Repository Service (GeoDaRRS) workshop,<sup>1</sup> held in August 2018, at NCAR. The workshop brought together more than 60 geoscience researchers, technology experts, scientific publishers, funders, and data repository personnel to discuss data management challenges and opportunities within the geosciences. This included exploring whether new services are needed to complement existing data facilities, particularly in the areas of 1) data management planning support resources and 2) repository services for geoscience researchers who have data that do not fit in any existing repository. More details on the workshop agenda and recommendations are available in the final workshop report (Mayernik et al. 2018).

<sup>1</sup> Funded by the National Science Foundation.

**AFFILIATIONS:** SCHUSTER, MAYERNIK, HOU, AND STOSSMEISTER—National Center for Atmospheric Research, Boulder, Colorado; DOWNS—Center for International Earth Science Information Network, Columbia University, Palisades, New York; KINKADE—Woods Hole Oceanographic Institution, Woods Hole, Massachusetts; NGUYEN—Department of Chemistry, University of California, Irvine, Irvine, California; RAMAMURTHY—University Corporation for Atmospheric Research, Boulder, Colorado; ZHANG—Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania  
**CORRESPONDING AUTHOR:** Douglas C. Schuster, [schuster@ucar.edu](mailto:schuster@ucar.edu)

DOI:10.1175/BAMS-D-18-0319.1

In final form 19 February 2019

©2019 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

## **DATA MANAGEMENT CHALLENGES AND RECOMMENDATIONS.**

The GeoDaRRS presentations and breakout discussions touched on a number of issues related to data sharing that have existed within scientific institutions for hundreds of years. Many salient disincentives to data sharing are well known, such as the potential for researchers to be “scooped” on important scientific findings; the possibility for data to be misunderstood by secondary users; the lack of clear reward structures for data sharing; and the time and effort required to clean, document, and provide access and support to data (Fienberg et al. 1985; Borgman 2012; Tenopir et al. 2011).

On the data repository side, repositories face practical constraints related to technical, human, and financial factors. Data storage requirements continue to grow, and repositories must continually refresh their software and hardware infrastructures. Repositories must also build staff expertise to support specific kinds of data and data systems. Many geoscience data repositories are funded via short-term grants or cooperative agreements that are subject to competitive bidding on regular intervals, which makes planning, expanding, and sequencing data repository operations significant challenges.

Workshop discussions focused on addressing related challenges across seven topic areas: 1) long-term, scalable data curation; 2) education and training; 3) data management plans; 4) funder and publisher policies; 5) strategic partnerships; 6) legacy data; and 7) tools and services. Below we summarize recommendations that emerged from the workshop.

### **LONG-TERM, SCALABLE DATA CURATION.**

**Challenges.** Researchers are now required to document, archive, and cite data by funders and publishers to support open access requirements. Many researchers do not know what repositories are available to meet these requirements. Second, projects that generate large data volumes, such as model outputs, face additional data storage and archiving challenges. At present, individuals dealing with these issues develop temporary, ad hoc solutions, including putting data on cloud infrastructures, local research group servers, and university-operated servers. In many cases these ad hoc solutions are not sustainable after short-term grants end.

**Recommendations.** Long-term support for the data curation needs of the geoscience community is critical for providing truly open access to data. Several different approaches to providing sustainable open

access to data are possible. These include 1) augment existing geoscience data repositories to scale up their capacity, 2) identify nonspecialized data repositories that fulfill open access objectives, 3) develop a data repository liaison service, and 4) create new data repository services. Further investigation will be needed to understand the relative merits and drawbacks of each approach for current and future needs. Ultimately, guidance regarding the use of these approaches may involve some combination of all four approaches, perhaps in a distributed model where organizations share expertise and/or infrastructure.

### **EDUCATION AND TRAINING.**

**Challenges.** Many problems that we encounter today when dealing with data result from people not being aware of data and data management practices in order to include best practices as part of their research workflows. Also, few formal educational programs have been established for developing professionals who are knowledgeable in scientific computing and data. As a result, scientific institutions have difficulty hiring people with the skill sets required to work on new technologies, such as cloud-based systems.

**Recommendations.** There is a need for programs to better support data management education within scientific, computing, information, and data disciplines. These programs will need to be designed appropriately to engage and involve scientists who face significant demands on their time.

Data management training and resources for researchers need to be improved and better publicized. For scientists, training on data management best practices and writing good data management plans is essential to engender thinking about data management from the beginning of projects.

### **DATA MANAGEMENT PLANS.**

**Challenges.** There is a perception in the scientific community that data management plans (DMPs) are an afterthought and that there is little accountability for following through on a data management plan. Additionally, although there are numerous templates available for data management plans, insufficient guidance exists within the geoscience community on which templates to use. Finally, repositories have difficulty in resourcing for incoming data products when not included at the project planning phase.

**Recommendations.** Grant proposal reviewers should 1) review data management planning according to multiple explicit criteria including sufficiency, resourcing,



and execution plans and 2) scrutinize this section as critically as all other sections of a proposal. Funders should also provide well defined data management guidelines and expectations for potential grantees that are not overly burdensome.

An efficient mechanism for grantees to update and comment on their DMPs during the annual reports would help improve the accountability for the DMPs. Funding agencies already ask grantees to report on a project's progress on a yearly basis. These yearly report submissions could provide a periodic way for grantees to report on their data management activities. However, such reporting requirements should not put too much of a time burden on researchers.

Data repositories need to be brought into the DMP conversations in the initial stages of the project planning process. When researchers know specifically where they plan to deposit data, they are much more likely to follow through with their plan. Data repositories can also provide specific examples and details that can inform project data management tasks and costs.

The geoscience community should foster a DMP tool ecosystem. Proposal writers should be encouraged to use existing DMP tools, and the community should develop and implement new tools to support, streamline, and bring consistency to DMP development.

## **FUNDER AND PUBLISHER POLICIES.**

**Challenges.** Inconsistent data management policies across funders and publishers create a complex landscape for researchers to navigate. This can lead to challenges in publishing research findings, and fulfilling funder data management requirements. Particularly, it may be impractical for projects that generate large data volumes, such as model outputs, to retain the full data record to support open access requirements.

**Recommendations.** All stakeholders should be clear on the core drivers and principles that motivate their data policies. For example, ensuring reproducibility may involve different work than ensuring data accessibility. Roles and responsibilities associated with these drivers and principles should likewise be defined clearly. More clarity on these points should help with consistency in data policy implementations across stakeholders.

Improved coordination across all stakeholders is needed to bring consistency to the data policy landscape. By contributing efforts toward data policy

coordination, stakeholders can demonstrate their commitment to bring about positive change in the data policy landscape.

Specific scientific research communities need to discuss and formalize data retention guidelines. This was identified during the GeoDaRRS workshop as a particular need for the atmospheric and hydrological modeling communities, where saving all components of large data volume model outputs for an extended duration may be impractical and unnecessary.

## **STRATEGIC PARTNERSHIPS. Challenges.**

Many federal agencies, universities, and research institutions face similar data management challenges yet build disconnected, insufficient, and/or independent solutions. This can lead to a duplication of effort and spending. Additionally, the purchasing of resources such as computing, storage, and cloud-based solutions can be prohibitively expensive for individual projects or institutions.

**Recommendations.** Strategic partnerships across federal agencies could reduce costs through shared data storage and curation services. For example, various federal agencies could work together to discuss common data infrastructure needs to facilitate potential cost sharing for colocated storage and cloud computing services.

Strategies need to be developed at the agency level to employ cloud computing and storage. For instance, a funding agency could execute bulk purchases of cloud computing resources, which could then be allocated to facilities and grantees through the proposal process.

## **LEGACY DATA. Challenges.**

Legacy data created/collected through past projects exist throughout the research community. In many cases, these data do not adhere to the format and metadata requirements needed to support long-term curation. When the grants used to create/collect these legacy data expire, researchers have few options for data cleaning, storage, or archiving.

**Recommendations.** Stakeholders should work together to provide researchers with clear paths to support curation and rescue of data collected via past projects. These might include 1) small grants to support researchers and/or data professionals to quality check, document, and reformat existing data in order to deposit them into a repository and 2) coordinated efforts to partner legacy data recovery initiatives with available data repositories.

**TOOLS AND SERVICES.** *Challenges.* Scientists push the boundaries of existing tools, often customizing or extending tools in unique ways. “Off the shelf” software is thus not always sufficient as a data analysis solution. Additionally, data storage costs continue to be an ongoing challenge. Despite the decrease in storage costs on a per-unit basis, total storage costs continue to increase as the amount of data generated through computational models and new observational instruments and sensors also increases.

*Recommendations.* All stakeholders should recognize the importance of open source software communities and contribute to these efforts where possible. Open source software has become instrumental in providing scientists with high-quality data analysis tools, along with communities of practice on how to best extend and contribute back to the same code bases.

Data repositories should investigate whether cost efficiencies can be gained by sharing data storage infrastructure. This could involve conceptual (and potentially financial) separation between the data storage function and the data curation function that repositories provide. This could also involve developing ways to adjust data storage tiers to accommodate variable costs in accordance with usage.

**CONCLUDING REMARKS.** The recommendations outlined in this meeting summary are intended to provide concrete steps on how stakeholders can move forward and work to address the many data

management challenges faced by the geoscience research community. Changes in the broader research culture will be needed to enhance the value of data management activities in research workflows and will require champions in disciplinary research communities to amplify the message and pass along their knowledge to colleagues.

**ACKNOWLEDGMENTS.** The National Science Foundation (NSF) provided the funding support for this workshop. We also thank Cecilia Banner and Elizabeth Faircloth of NCAR for administrative and logistical support.

## REFERENCES

- Borgman, C. L., 2012: The conundrum of sharing research data. *J. Amer. Soc. Inf. Sci. Technol.*, **63**, 1059–1078, <https://doi.org/10.1002/asi.22634>.
- Fienberg, S. E., M. E. Martin, and M. L. Straf, Eds., 1985: *Sharing Research Data*. National Academies Press, 225 pp.
- Mayernik, M. S., D. Schuster, S. Hou, and G. J. Stossmeister, 2018: Geoscience Digital Data Resource and Repository Service (GeoDaRRS) Workshop report. NCAR Tech. Note NCAR/TN-552+PROC, 43 pp., <https://doi.org/10.5065/D6NC601B>.
- Tenopir, C., S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame, 2011: Data sharing by scientists: Practices and perceptions. *PLOS ONE*, **6**, e21101, <https://doi.org/10.1371/journal.pone.0021101>.



# THANK YOU

MORE THAN  
21,000 TEACHERS  
DIRECTLY TAUGHT

OVER 175,000  
TEACHERS PEER  
TRAINED

MILLIONS OF  
STUDENTS  
BENEFIT

[ametsoc.org/educationprogram](http://ametsoc.org/educationprogram)

Thank you for supporting the AMS Education Program!  
Your support contributes to our ability to provide professional  
development programs and opportunities in weather, ocean  
and climate science to K-12 teachers.



# AMS BOOKS

RESEARCH APPLICATIONS HISTORY

AMS MEMBERS GET FREE

## CLIMATE

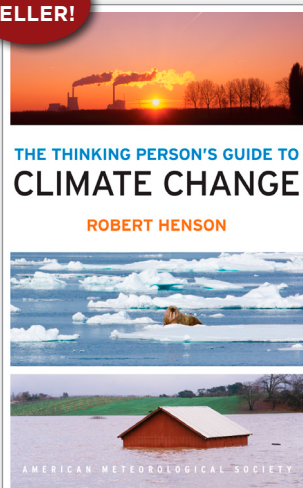
### The Thinking Person's Guide to Climate Change

ROBERT HENSON

This fully updated and expanded revision of *The Rough Guide to Climate Change* combines years of data with recent research. It is the most comprehensive overview of climate science, acknowledging controversies but standing strong in its stance that the climate is changing—and something needs to be done.

© 2014, PAPERBACK, 520 PAGES,  
ISBN: 978-1-935704-73-7  
LIST \$30 MEMBER \$20

BEST  
SELLER!

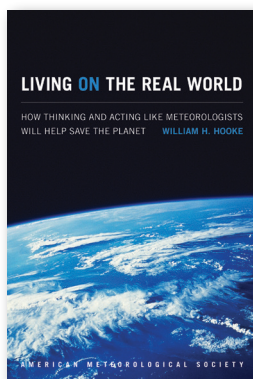


### Climate Conundrums: What the Climate Debate Reveals about Us

WILLIAM B. GAIL

This is a journey through how we think, individually and collectively, about humanity's relationship with nature, and more. Can we make nature better? Could science and religion reconcile? Gail's insights on such issues help us better understand who we are and find a way forward.

© 2014, PAPERBACK, 240 PAGES,  
ISBN: 978-1-935704-74-4 LIST \$30 MEMBER \$20



### Living on the Real World: How Thinking and Acting Like Meteorologists Will Help Save the Planet

WILLIAM H. HOOKE

Meteorologists focus on small bits of information while using frequent collaboration to make decisions. With climate change a reality, William H. Hooke suggests we look to the way meteorologists operate as a model for how we can solve the 21st century's most urgent environmental problems.

© 2014, PAPERBACK, 272 PAGES, ISBN 978-1-935704-56-0 LIST \$30 MEMBER \$22

### Synoptic-Dynamic Meteorology Lab Manual:

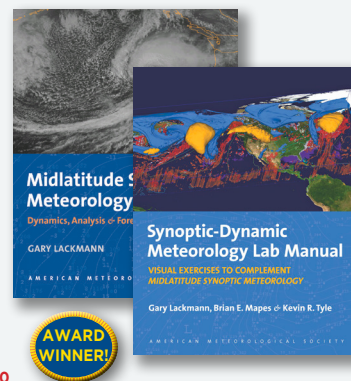
#### Visual Exercises to Complement Midlatitude Synoptic Meteorology

GARY LACKMANN,  
BRIAN E. MAPES, AND  
KEVIN R. TYLE

These labs link theoretical concepts with ground-breaking visualization to elucidate concepts taught in the award-winning companion textbook by Gary Lackmann, *Midlatitude Synoptic Meteorology*.

© 2017, PAPERBACK, 126 PAGES,  
ISBN 978-1-878220-26-4

LIST \$80 MEMBER \$60 STUDENT \$50

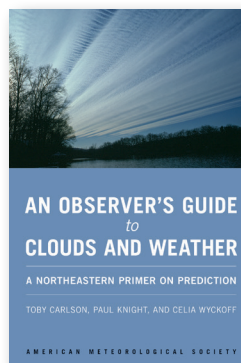


## GUIDES

### An Observer's Guide to Clouds and Weather:

#### A Northeastern Primer on Prediction

TOBY CARLSON, PAUL KNIGHT,  
AND CELIA WYCKOFF



With help from Penn State experts, start at the beginning and go deep. This primer, intended for both serious enthusiasts and new meteorology students, will leave you with both refined observation skills and an understanding of the complex science behind the weather: the ingredients for making reliable predictions of your own. It connects fundamental meteorological concepts with the processes that shape weather patterns, and will make an expert of any dedicated reader.

© 2014, PAPERBACK, 210 PAGES,  
ISBN: 978-1-935704-58-4 LIST \$30 MEMBER \$20

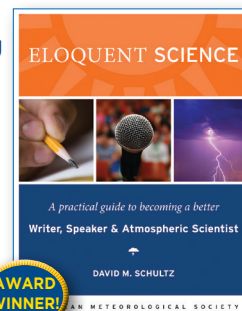
### Eloquent Science: A Practical Guide to Becoming a Better Writer, Speaker, and Atmospheric Scientist

DAVID M. SCHULTZ

The ultimate communications manual for undergraduate and graduate students as well as researchers in the atmospheric sciences and their intersecting disciplines.

© 2009, PAPERBACK, 440 PAGES,  
ISBN 978-1-878220-91-2

LIST \$45 MEMBER \$30



To order: [bookstore.ametsoc.org](http://bookstore.ametsoc.org), 617-226-3998, or use the order form in this magazine

## NEW

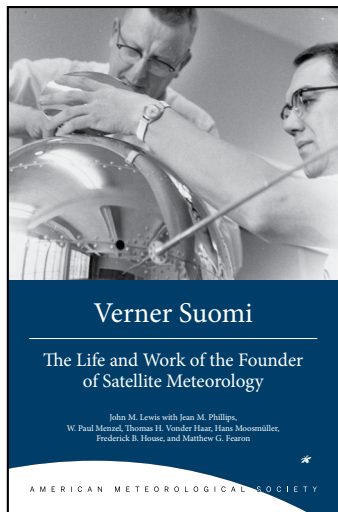
### Verner Suomi: The Life and Work of the Founder of Satellite Meteorology

JOHN M. LEWIS WITH  
JEAN M. PHILLIPS, W. PAUL  
MENZEL, THOMAS H. VONDER  
HAAR, HANS MOOSMÜLLER,  
FREDERICK B. HOUSE,  
AND MATTHEW G. FEARON

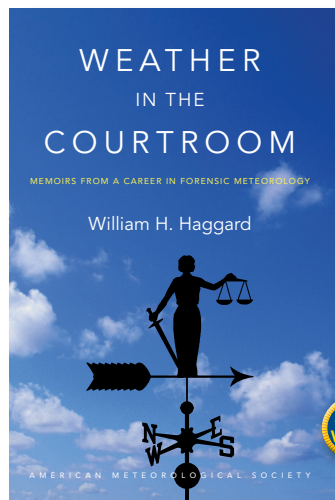
Born in a Minnesotan mining town, Suomi would spend his best years next door in Wisconsin, but not before seeing the whole world—from space, that is. This is the story of the scientist, inventor, and teacher who founded satellite meteorology, written by members of the communities that grew up around his groundbreaking work.

**LIST \$30 MEMBER \$20**

© 2016, PAPERBACK, 240 PAGES, ISBN: 978-1-944970-22-2



### Weather in the Courtroom: Memoirs from a Career iForensic Meteorology



WILLIAM H. HAGGARD

From a pioneering forensic meteorologist, the inside scoop on legendary litigations, including the disappearance of an Alaskan congressman's airplane in 1972, the collapse of Tampa Bay's Skyway Bridge in 1980, and the crash of Delta Flight 191 in Dallas/Fort Worth in 1985.

**LIST \$30 MEMBER \$20**

© 2016, PAPERBACK, 240 PAGES,  
ISBN: 978-1-940033-95-2



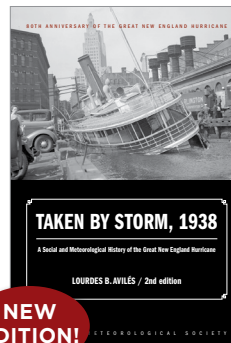
#### Booksellers, groups, or for examination copies:

The University of Chicago Press:  
1-800-621-2736 (US & Canada)  
773-702-7000 (all others)  
custserv@press.uchicago.edu



**AMS**  
American Meteorological Society

## HISTORY



**NEW  
EDITION!**

### Taken by Storm, 1938:

A Social and Meteorological History of the Great New England Hurricane, 2nd Ed.

LOURDES B. AVILÉS

80TH  
ANNIVERSARY  
OF STORM

The science behind the 1938 Hurricane, which hit New England unannounced, is presented here for the first time along with new data that

sheds light on the motivations of the Weather Bureau forecasters. This compelling history successfully weaves science, historical accounts, and social analyses to create a comprehensive picture of the most powerful and devastating hurricane to hit New England to date.

© 2018, PAPERBACK, 288 PAGES, ISBN: 978-1-944970-24-6

**LIST \$30 MEMBER \$20**

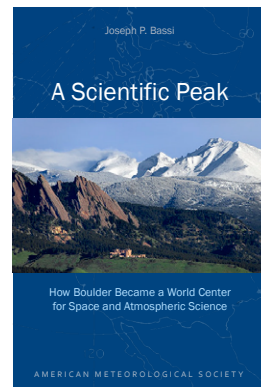
### A Scientific Peak: How Boulder Became a World Center for Space and Atmospheric Science

JOSEPH P. BASSI

How did big science come to Boulder, Colorado? Joe Bassi introduces us to the characters, including Harvard sun-Earth researcher Walter Orr Roberts, and the unexpected brew of politics, passion, and sheer luck that during the Cold War era transformed this "Scientific Siberia" to home of NCAR and NOAA.

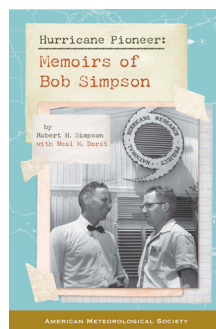
© 2015, PAPERBACK, 264 PAGES, ISBN: 978-1-935704-85-0

**LIST PRICE: \$35.00 MEMBER PRICE: \$25.00**



### Hurricane Pioneer: Memoirs of Bob Simpson

ROBERT H. SIMPSON AND NEAL DORST



In 1951, Bob Simpson rode a plane into a hurricane—just one of the many pioneering exploits you'll find in these memoirs. Bob and his wife Joanne are meteorological icons: Bob was the first director of the National Hurricane Research Project and a director of the National Hurricane Center. He helped to create the Saffir-Simpson Hurricane Scale; the

public knows well his Categories 1-5. Proceeds from this book help support the AMS's K. Vic Ooyama Scholarship Fund.

© 2015, PAPERBACK, 156 PAGES

ISBN: 978-1-935704-75-1

**LIST \$25**

**MEMBER \$20**

# AMS Publications introduces



## A free online tool for authors!

Communicate Your Research More Effectively

Increase Your Work's Impact

- Maximize citations and downloads of your article
- Open up your research to new audiences
- Explain and share your work in only 10 minutes
- Increase full-text article downloads by 23%
- Access publication metrics via the author dashboard

Learn more about using Kudos with AMS

<https://www.ametsoc.org/kudos/>